

Application of Independent Component Analysis Techniques in Speckle Noise Reduction of Retinal OCT Images

Ahmadreza Baghaie¹, Roshan M. D'Souza², Zeyun Yu³

¹*Department of Electrical Engineering,* ²*Department of Mechanical Engineering,*
³*Department of Computer Science,*
University of Wisconsin-Milwaukee, Milwaukee, WI, USA

Abstract

Optical Coherence Tomography (OCT) is an emerging technique in the field of biomedical imaging, with applications in ophthalmology, dermatology, coronary imaging etc. OCT images usually suffer from a granular pattern, called speckle noise, which restricts the process of interpretation. Therefore the need for speckle noise reduction techniques is of high importance. To the best of our knowledge, use of Independent Component Analysis (ICA) techniques has never been explored for speckle reduction of OCT images. Here, a comparative study of several ICA techniques (InfoMax, JADE, FastICA and SOBI) is provided for noise reduction of retinal OCT images. Having multiple B-scans of the same location, the eye movements are compensated using a rigid registration technique. Then, different ICA techniques are applied to the aggregated set of B-scans for extracting the noise-free image. Signal-to-Noise-Ratio (SNR), Contrast-to-Noise-Ratio (CNR) and Equivalent-Number-of-Looks (ENL), as well as analysis on the computational complexity of the methods, are considered as metrics for comparison. The results show that use of ICA can be beneficial, especially in case of having fewer number of B-scans.

Keywords: Independent Component Analysis, Speckle Reduction, Optical Coherence Tomography (OCT)

1. Introduction

OCT is a powerful imaging system for non-invasive acquisition of 3D volumetric images of tissues [1], with applications in ophthalmology, dermatology, coronary imaging etc. Due to its underlying physics, which is common in narrow-band detection systems like Synthetic-Aperture Radar (SAR) and ultrasound, OCT images usually suffer from a granular pattern called *speckle*. Not only the optical properties of the system, but also the motion of the subject to be imaged, size and temporal coherence of the light source, multiple scattering, phase deviation of the beam and aperture of the detector can affect the speckle [2]. Fig. 1 shows a sample retinal OCT image, highly degraded by speckle noise.

Speckle is considered to be multiplicative noise, in contrast to the additive Gaussian noise. Limited dynamic range of displays requires us to compress the OCT signals usually by a logarithmic transform, which converts the multiplicative speckle noise to additive noise [3]. Two major classes of speckle noise reduction techniques are: 1) methods of noise reduction during the acquisition time and 2) post-processing techniques. In the first class multiple uncorrelated recordings are averaged. This includes spatial [4], angular [5], polarization [6] and frequency [7] compounding techniques. As for post-processing, anisotropic diffusion-based techniques [3] and multi-scale/multi-resolution geometric representation techniques [8] are of high interest between scholars. Use of compressive sensing and sparse representation have also been explored in the past few years [9]. For a more complete review on the different image analysis techniques in OCT image processing, including noise reduction, the reader is referred to [10] and references therein.

Post-processing averaging/median filtering is also an interesting method for speckle reduction. In such techniques, multiple B-scans of the same location are acquired and then the average/median is taken. The misalignment between the different B-scans is usually compensated with a parametric image registration technique. Theoretically, having N B-scans with uncorrelated speckle, SNR can be improved by a factor of \sqrt{N} . The works presented in [11, 12] can be mentioned as examples. Recently the use of sparse and low-rank decomposition based batch image alignment was explored by the authors [13].

In this paper the use of Independent Component Analysis (ICA) techniques for speckle noise reduction of retinal OCT images is explored, which to the best of our knowledge has never been investigated before. Having

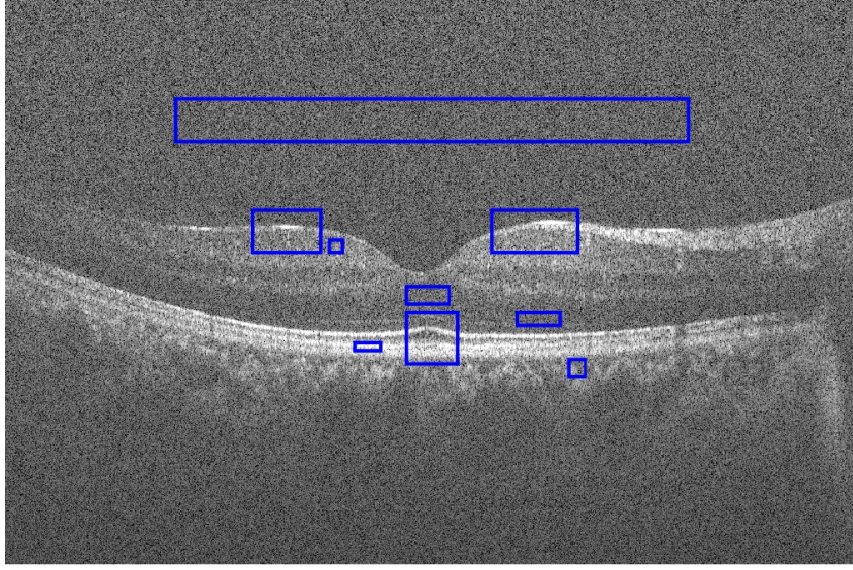


Figure 1: Sample retinal OCT image degraded by speckle noise; selected ROIs are shown with blue rectangles.

multiple B-scans of the same location in retina, the eye movement is compensated by considering a rigid transformation between consecutive B-scans using ImageJ [14]. Having negligible eye motion within each B-scan, the need for deformable registration techniques [15, 16] can be eliminated. Then, several ICA techniques are used for extracting the noise-free image from multiple noisy B-scans. SNR, CNR and ENL are considered as metrics for comparing the performance of different methods. Investigating the results reveals interesting facts regarding the impact of using ICA for speckle noise reduction. Especially in case of having a few number of images in which ICA techniques provide better performance in comparison to median filtering. But increasing the number of images causes the ICA techniques to perform poorly, since the main assumption in ICA is having non-Gaussian uncorrelated components which can't be completely satisfied here.

2. Independent Component Analysis (ICA)

ICA is one of the most widely used techniques for Blind Source Separation (BSS) [17]. The problem of BSS consists of having interfering signals from multiple sources recorded and trying to find the individual source signals from

these mixed recordings. The very well-known example is a cocktail party, with multiple people talking while there are recorders in different places of the room. This can be mathematically modeled by considering $s_i(t), i = 1, \dots, N$ as the set of sources and $x_i(t), i = 1, \dots, N$ as the observations. Therefore:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (1)$$

where \mathbf{A} is the unknown *mixing matrix*. Recovering the original signals $s_i(t)$ from the recordings $x_i(t)$ is the objective of BSS. Therefore, estimating the unmixing matrix \mathbf{W} such that $\mathbf{W} = \mathbf{A}^{-1}$, the sources can be estimated:

$$\hat{\mathbf{s}}(t) = \mathbf{W}\mathbf{x}(t) \quad (2)$$

But how to find \mathbf{W} when you are "blind" to the nature of mixing matrix and the source signals themselves? This question results in defining a set of constraints on the nature of source signals. As is obvious from the name, ICA, the components should be *statistically independent*. Mathematically speaking, having two random variables s_1 and s_2 , the joint probability density function (pdf) $p_{1,2}(s_1, s_2)$ for independent variables can be computed as:

$$p_{1,2}(s_1, s_2) = p_1(s_1)p_2(s_2) \quad (3)$$

where $p_1(s_1)$ and $p_2(s_2)$ are the marginal pdf's.

Another constraint of the signal sources is that they should be non-Gaussian. Based on the central limit theorem, the distribution of a sum of independent signals with arbitrary distributions tends to be a Gaussian distribution, which basically means any Gaussian distribution can be considered as a sum of multiple independent signals. Therefore having Gaussian signals as components prevents us from being able to distinguish them properly.

ICA has two ambiguities. The variance and the sign of the independent components cannot be estimated. This is usually insignificant in most of the applications. The second ambiguity is regarding the order of the extracted components which cannot be specified.

As for pre-processing the data before independent component analysis, centering the data (subtracting the mean value) is usually the first step in order to make the ICA procedure easier. The mean will be added after extracting the components. Whitening the data for having uncorrelated components with unit variances is next.

Here four well-known ICA methods are used for noise reduction of retinal OCT images: InfoMax (RUNICA) [18], FastICA [19], JADE [20] and SOBI [21]. In each one of these techniques a contrast/likelihood function is defined and optimized for achieving the optimal separation between sources. The InfoMax method works based on maximizing the output entropy (information flow) of a neural network with non-linear outputs. FastICA on the other hand tries to maximize the non-Gaussianity using an approximation for negentropy utilizing a fixed-point numerical scheme. Joint Approximate Diagonalisation of Eigen-matrices (JADE) method works based on diagonalization of the sample data's fourth-order cumulants of the whitened process created from the sample data and a whitening matrix. On the other hand, Second Order Blind Identification (SOBI) assumes that the sample data is gathered from a set of temporally correlated sources and tries to separate them by joint diagonalization of several correlation matrices. These techniques are widely used in different applications, like electroencephalogram (EEG) source separation [22]. In the following section brief overviews will be given for each one of the above-mentioned methods. The reader is referred to [17, 23, 24, 25, 26, 27, 28, 29] and references therein, as well as EEGLAB [30] and ICA Central (<http://perso.telecom-paristech.fr/~cardoso/icacentral/>) for further reading regarding other ICA techniques and pointers for implementations.

3. Methods

3.1. InfoMax

InfoMax is based on a neural network approach [18], which tries to maximize the mutual information of the network's output Y about its input X , as follows:

$$I(Y, X) = H(Y) - H(Y|X) \quad (4)$$

where $H(Y)$ is the entropy of output Y while $H(Y|X)$ represents the entropy of the output that is not a result of input. Considering only the *gradient* of information theoretic quantities with respect to some parameter w in the network, while knowing that $H(Y|X)$ does not depend on w , we can have:

$$\frac{\partial}{\partial w} I(X, Y) = \frac{\partial}{\partial w} H(Y) \quad (5)$$

Having a neural network with input vector \mathbf{x} , weight matrix \mathbf{W} , bias vector \mathbf{w}_0 and monotonically transformed output vector $\mathbf{y} = g(\mathbf{W}\mathbf{x}) + \mathbf{w}_0$

with $g(u)$ being a sigmoidal function ($g(u) = (1 + e^{-u})^{-1}$), the learning rule will be of the form:

$$\Delta \mathbf{W} \propto [\mathbf{W}^T]^{-1} + (\mathbf{1} - 2\mathbf{y})\mathbf{x}^T \quad (6)$$

where $\mathbf{1}$ is a vector of ones. At each iteration \mathbf{W} is updated until it reaches convergence.

3.2. FastICA

The differential entropy H of a random vector y with density function $f(y)$ can be defined as:

$$H(y) = - \int f(y) \log f(y) dy \quad (7)$$

It has been proved that [31, 32] a Gaussian variable has the largest entropy among all random variables of equal variance. Therefore to obtain a measure for non-Gaussianity, a modified definition of differential entropy is used which is called *negentropy*:

$$J(y) = H(y_{gauss}) - H(y) \quad (8)$$

in which $H(y_{gauss})$ is a Gaussian random variable of the same covariance matrix as y . This measure is always non-negative since the second term in subtraction is always smaller or equal to the first term. For FastICA, an approximation of the negentropy is used:

$$J(y) \propto [E\{G(y)\} - E\{G(v)\}]^2 \quad (9)$$

v being a zero-mean and unit variance Gaussian variable and G a non-quadratic function. Two examples for proper function G are $G_1(u) = \frac{1}{a_1} \log(\cosh(a_1 u))$ and $G_2(u) = -\exp(-u^2/2)$ for $1 \leq a_1 \leq 2$. More elaboration on these choices can be found in [26]. Maximization of the non-Gaussianity of $w^T x$ is done using a fixed-point scheme consisting of several steps, with g being the derivative of G , as follows:

1. initialization of the weight vector w ;
2. $w^+ = E\{x(g(w^T x))\} - E\{g(w^T x)\}w$;
3. $w = w^+ / \|w^+\|$;
4. **goto** step 2, until convergence.

In this formulation, the convergence happens when the new and old w vectors are in the same direction which means having their dot product equal to zero. Of course this can happen even if they are in opposite directions too, which is valid since that represents one of the ambiguities of the ICA.

3.3. SOBI

Assuming a more general case of mixed sources with addition of noise $\mathbf{n}(t)$ we have:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \quad (10)$$

where \mathbf{A} is a full-ranked complex matrix. The main assumption is to have second order stationary and mutually uncorrelated sources while the additive noise is assumed to be spatially and temporally white and uncorrelated with the source signals. SOBI tries to separate the components by joint diagonalization of several correlation matrices. Based on these assumptions, the algorithm of robust SOBI for N samples consists of several steps as follows [29]:

1. robust orthogonalization of the sensor signals, $\bar{x} = \mathbf{Q}x(t)$, \mathbf{Q} being the orthogonalization matrix;
2. estimate the set of covariance matrices:

$$\hat{\mathbf{R}}_{\bar{x}}(p_i) = \frac{1}{N} \sum_{k=1}^N \bar{x}(t) \bar{x}^T(t - p_i) = \mathbf{Q} \hat{\mathbf{R}}_x(p_i) \mathbf{Q}^T$$

for a set of time lags (p_1, p_2, \dots, p_L) ;

3. performing joint approximate diagonalization: $\mathbf{R}_{\bar{x}}(p_i) = \mathbf{U} \mathbf{D}_i \mathbf{U}^T, \forall i$, which means estimating the orthogonal matrix \mathbf{U} ;
4. estimating the source signals

$$\hat{s}(t) = \mathbf{U}^T \mathbf{Q} x(t)$$

and the mixing matrix

$$\hat{\mathbf{W}} = \mathbf{Q}^+ \mathbf{U}.$$

3.4. JADE

The JADE (Joint Approximate Diagonalization of Eigenmatrices) is a natural extension of the SOBI. The algorithm consists of the following steps [29]:

1. robust pre-whitening or orthogonalization: $\bar{x} = \mathbf{Q}x(t)$;
2. performing eigen value decomposition (EVD) of the sampled contracted quadri-covariance matrix

$$\begin{aligned} \mathbf{C}_{\bar{x}}(\mathbf{I}) &= \frac{1}{N} \sum_{k=1}^N \{\bar{x}^T(t)\bar{x}(t)\bar{x}(t)\bar{x}^T(t)\} - 2\hat{\mathbf{R}}_{\bar{x}}(0)\hat{\mathbf{R}}_{\bar{x}}(0) - tr(\hat{\mathbf{R}}_{\bar{x}}(0))\hat{\mathbf{R}}_{\bar{x}}(0) \\ &= \hat{\mathbf{U}}\mathbf{\Lambda}_I\hat{\mathbf{U}}^T \end{aligned} \quad (11)$$

with $\hat{\mathbf{R}}_{\bar{x}}(0) = \frac{1}{N} \sum_{k=1}^N \{\bar{x}(t)\bar{x}^T(t)\}$ and $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \dots, \hat{\mathbf{u}}_n]$;

3. estimating the n sampled contracted quadri-covariance matrices:

$$\begin{aligned} \mathbf{C}_{\bar{x}}(\mathbf{E}_p) &= \frac{1}{N} \sum_{k=1}^N \{\bar{x}^T(t)\mathbf{E}_p\bar{x}(t)\bar{x}(t)\bar{x}^T(t)\} - \hat{\mathbf{R}}_{\bar{x}}(0)\mathbf{E}_p\hat{\mathbf{R}}_{\bar{x}}(0) \\ &\quad - tr(\mathbf{E}_p\hat{\mathbf{R}}_{\bar{x}}(0))\hat{\mathbf{R}}_{\bar{x}}(0) - \hat{\mathbf{R}}_{\bar{x}}\mathbf{E}_p^T\hat{\mathbf{R}}_{\bar{x}} \end{aligned} \quad (12)$$

for $\mathbf{E}_p = \hat{\mathbf{u}}_p\hat{\mathbf{u}}_p^T, p = 1, 2, \dots, n$;

4. finding an orthogonal joint diagonalization matrix \mathbf{U} for all n matrices;
5. estimating the mixing matrix using $\hat{\mathbf{W}} = \mathbf{Q}^+\mathbf{U}$.

For more elaboration on the methods for joint approximate diagonalization, the reader is referred to [29].

4. Results and Discussion

For testing the methods, multiple B-scans of the same cross-section in retina are acquired using a Bioptigen OCT device (Bioptigen, Inc. Durham, NC 27709) available at AOIP (Advanced Ocular Imaging Program, Medical College of Wisconsin, Milwaukee, WI). Even though the subject is asked to fixate, the involuntary eye movements (tremor, drifts, micro-saccades) cause misalignment between consecutive B-scans. Considering the effect of eye movements within each B-scan as negligible, a parametric registration technique, namely rigid registration, is suitable for compensating the motion between B-scans. For this, the rigid registration technique implemented in ImageJ [14] is used. After aligning the B-scans, they are vectorized and aggregated into a data matrix and then ICA techniques, implemented in

MATLAB (The MathWorks, Inc, Natick, MA 01760), are applied to extract the independent components in the data matrix. The main component is the noise-free image, while the rest are only containing noise.

For assessing the performance of the methods, several metrics are considered. Considering 9 regions of interest (ROI) as depicted in Fig. 1 in the final result, one only containing background noise and the rest containing image features and homogeneous regions, the metrics can be defined as follows:

$$\begin{aligned} SNR_m &= 20 \times \log_{10}\left(\frac{\mu_m}{\sigma_b}\right) \\ CNR_m &= \frac{\mu_m - \mu_b}{\sqrt{\sigma_m^2 + \sigma_b^2}} \\ ENL_m &= \frac{\mu_m^2}{\sigma_m^2} \end{aligned} \quad (13)$$

where μ_b and σ_b are the mean and standard deviation of the background noise and μ_m and σ_m are the mean and standard deviation of the m -th ROI containing image features. ENL provides a measure for smoothness of homogeneous regions while SNR and CNR provide information on the signal and contrast of image features with respect to the background noise. The average of these metrics are considered here for comparison. Different numbers of images of the human retina in the central foveal region are considered for assessing the performance of the proposed algorithm.

Fig. 2 displays the improvement of SNR for different methods using different numbers of input B-scans. Theoretically, having N images, the SNR can be improved by \sqrt{N} using average/median filtering. As is obvious, this is not the case for ICA techniques. While all of the ICA techniques outperform median filtering at first, for using up to 20 input B-scans, only the SOBI remains superior for the rest and RUNICA, JADE and FastICA perform poorly. As mentioned before, one of the main constraints on the input data for being able to use ICA is having uncorrelated non-Gaussian components. This can't be satisfied, especially when using excessive number of images. The same pattern can be seen in Fig. 3 and Fig. 4 for the improvement of CNR and ENL, respectively. In case of SOBI, the assumption of having temporally correlated sources in the design of the algorithm makes it better in terms of SNR, CNR and ENL since the input images are acquired from the same location in retina, with the possibility of misalignment due to eye movement.

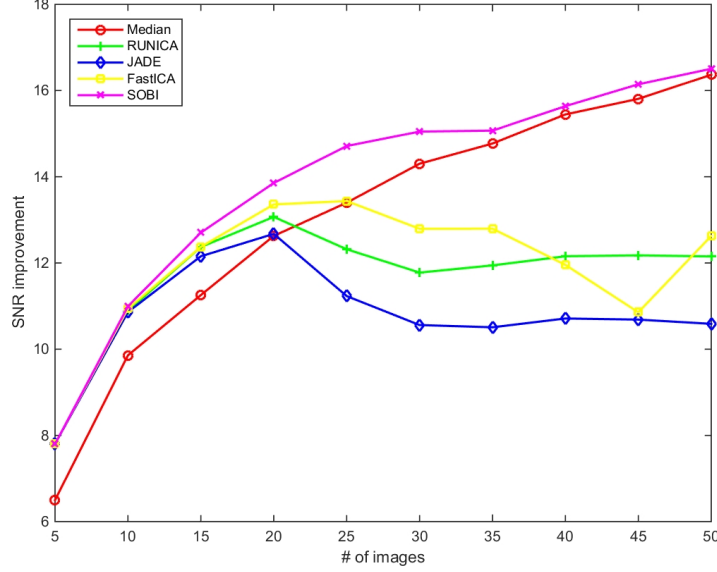


Figure 2: SNR improvement for different number of input images (5-50).

As for the computational complexity, while JADE shows a quadratic behavior for convergence, the rest show linear increment when having more input B-scans. Fig. 5 shows the graphs for computational time of different methods for different number of input images. Fig. 6 displays close-ups of the original and filtered version of the input images using different techniques for comparison.

5. Conclusion

In this paper, a new application of ICA techniques for noise reduction of retinal OCT images is proposed. Having a set of B-scans from the same location (to some extent) in retina, the eye movement is compensated using a parametric registration technique for compensating translation and rotation between consecutive B-scans. Then, taking advantage of four ICA techniques, RUNICA, JADE, FastICA and SOBI, the aggregated dataset is analyzed and the noise free image is extracted. Having N B-scans with uncorrelated speckle, the expected improvement in SNR for average/median filtering is \sqrt{N} . While this is satisfied for up to 20 input B-scans, with ICA techniques outperforming the median filtering, having more input B-scans results in poor performance in all of the ICA methods, except for SOBI.

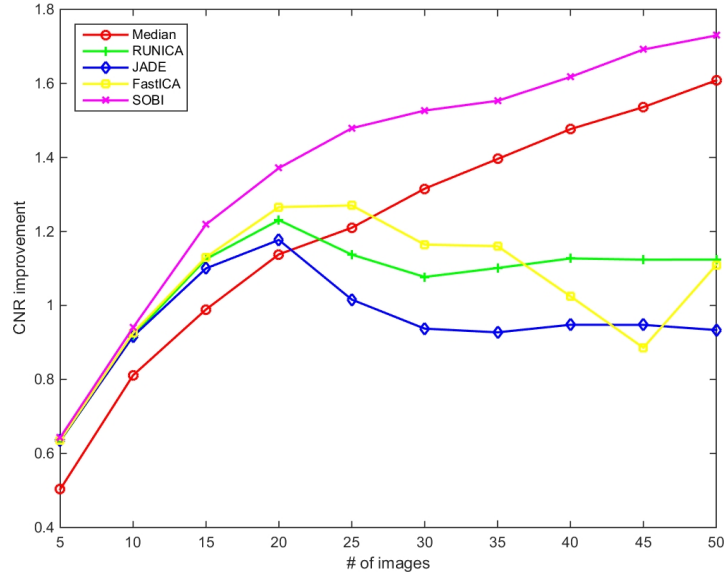


Figure 3: CNR improvement for different number of input images (5-50).

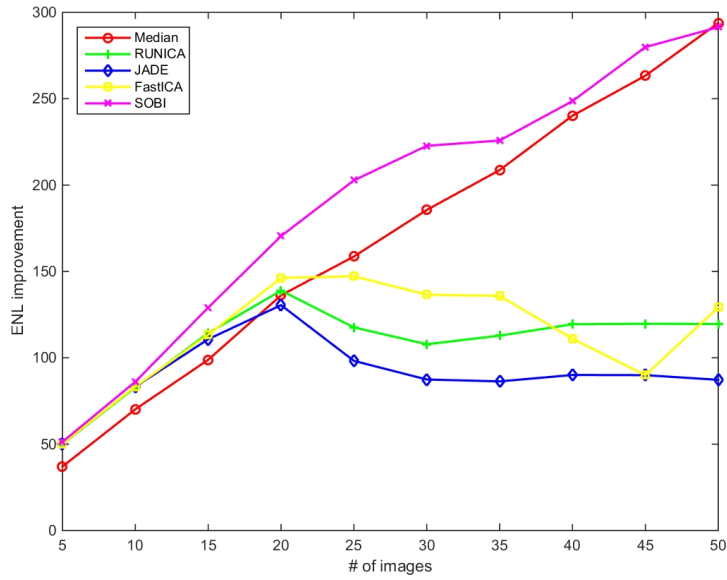


Figure 4: ENL improvement for different number of input images (5-50).

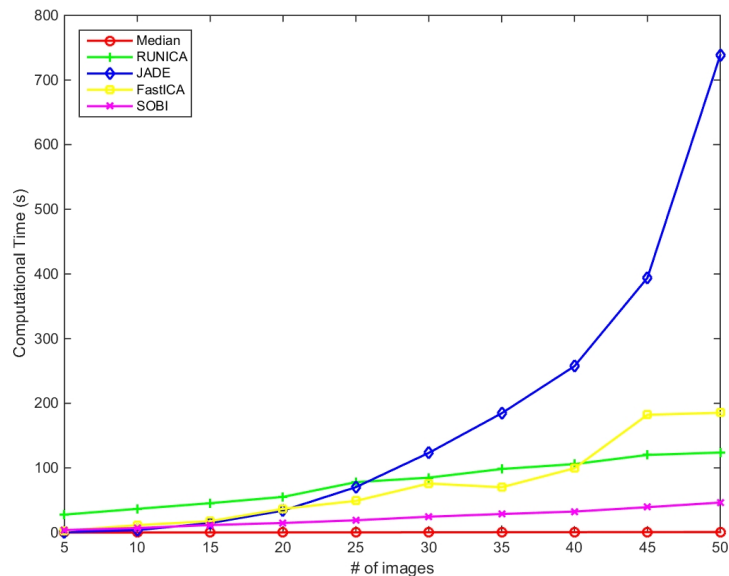


Figure 5: Comparison of computational times.

As for the comparison of the computational time, JADE has a quadratic behavior, while the rest show linear increment in computational time when having more images. Overall, SOBI is the best among the ICA techniques considered here in terms of performance based on SNR, CNR and ENL, while needing less computational power. As for pointers towards next possible areas for research, newer techniques for ICA/BSS as well as more analysis on the speckle noise model in OCT images should be explored and investigated, while reducing the computational complexity and exploring the possibility of using the techniques in an on-line manner can be considered as next steps.

6. Acknowledgment

The authors would like to thank Dr. Joseph Carroll from Advanced Ocular Imaging Program (AOIP), Medical College of Wisconsin (MCW), Milwaukee, WI for providing the data and insight. This work was partially supported by NIH P30EY001931. Support received by grant 8UL1TR000055 from the Clinical and Translational Science Award (CTSA) program of the National Center for Research Resources and the National Center for Advancing Translational Sciences. Support received by the Clinical and Translational Science Institute of Southeast Wisconsin through the Advancing a

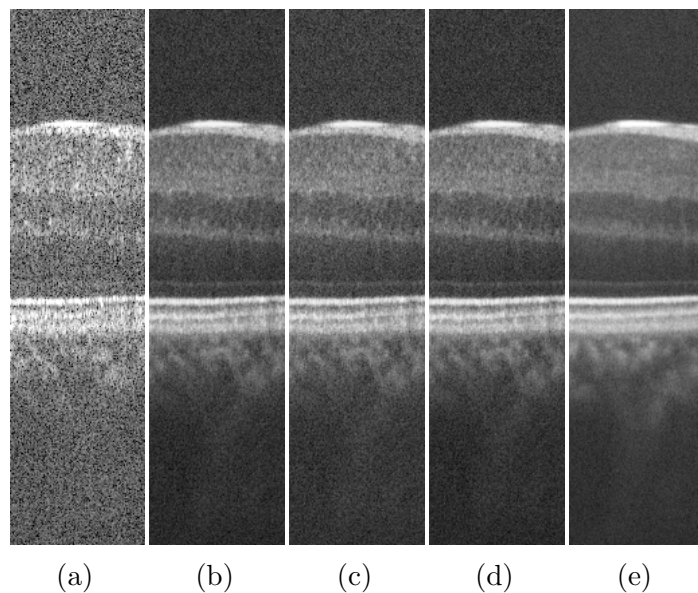


Figure 6: A portion of one of the input images (a) and the results of InfoMax (b), JADE (c), FastICA (d) and SOBI (e), resulted from 50 B-scans.

Healthier Wisconsin endowment of the Medical College of Wisconsin.

7. References

References

- [1] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, et al., Optical coherence tomography, *Science* 254 (1991) 1178–1181.
- [2] J. M. Schmitt, S. Xiang, K. M. Yung, Speckle in optical coherence tomography, *Journal of biomedical optics* 4 (1999) 95–105.
- [3] H. M. Salinas, D. C. Fernández, Comparison of pde-based nonlinear diffusion approaches for image enhancement and denoising in optical coherence tomography, *Medical Imaging, IEEE Transactions on* 26 (2007) 761–771.
- [4] M. R. Avanaki, R. Cernat, P. J. Tadrous, T. Tatla, A. G. Podoleanu, S. A. Hojjatoleslami, Spatial compounding algorithm for speckle reduc-

- tion of dynamic focus oct images, *Photonics Technology Letters, IEEE* 25 (2013) 1439–1442.
- [5] J. Schmitt, Array detection for speckle reduction in optical coherence microscopy, *Physics in Medicine and Biology* 42 (1997) 1427.
 - [6] M. Kobayashi, H. Hanafusa, K. Takada, J. Noda, Polarization-independent interferometric optical-time-domain reflectometer, *Light-wave Technology, Journal of* 9 (1991) 623–628.
 - [7] M. Pircher, E. Go, R. Leitgeb, A. F. Fercher, C. K. Hitzenberger, et al., Speckle reduction in optical coherence tomography by frequency compounding, *Journal of Biomedical Optics* 8 (2003) 565–569.
 - [8] A. Pizurica, L. Jovanov, B. Huysmans, V. Zlokolic, P. De Keyser, F. Dhaenens, W. Philips, Multiresolution denoising for optical coherence tomography: a review and evaluation, *Current Medical Imaging Reviews* 4 (2008) 270–284.
 - [9] L. Fang, S. Li, R. P. McNabb, Q. Nie, A. N. Kuo, C. Toth, J. Izatt, S. Farsiu, et al., Fast acquisition and reconstruction of optical coherence tomography images via sparse representation, *Medical Imaging, IEEE Transactions on* 32 (2013) 2034–2049.
 - [10] A. Baghaie, R. M. D’Souza, Z. Yu, State-of-the-art in retinal optical coherence tomography image analysis, *Quantitative Imaging in Medicine and Surgery* (2015).
 - [11] T. M. Jørgensen, J. Thomadsen, U. Christensen, W. Soliman, B. Sander, Enhancing the signal-to-noise ratio in ophthalmic optical coherence tomography by image registration method and clinical examples, *Journal of biomedical optics* 12 (2007) 041208–041208.
 - [12] D. Alonso-Caneiro, S. A. Read, M. J. Collins, Speckle reduction in optical coherence tomography imaging by affine-motion image registration, *Journal of biomedical optics* 16 (2011) 116027–1160275.
 - [13] A. Baghaie, R. M. D’souza, Z. Yu, Sparse and low rank decomposition based batch image alignment for speckle reduction of retinal oct images, *arXiv preprint arXiv:1411.4033* (2014).

- [14] C. A. Schneider, W. S. Rasband, K. W. Eliceiri, Nih image to imagej: 25 years of image analysis, *Nature methods* 9 (2012) 671–675.
- [15] A. Baghaie, Z. Yu, R. M. D’souza, Fast mesh-based medical image registration, in: *Advances in Visual Computing*, Springer, 2014, pp. 1–10.
- [16] A. Baghaie, Z. Yu, Curvature-based registration for slice interpolation of medical images, in: *Computational Modeling of Objects Presented in Images. Fundamentals, Methods, and Applications*, Springer, 2014, pp. 69–80.
- [17] G. R. Naik, D. K. Kumar, An overview of independent component analysis and its applications, *Informatica* 35 (2011).
- [18] A. J. Bell, T. J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, *Neural computation* 7 (1995) 1129–1159.
- [19] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, *Neural Networks, IEEE Transactions on* 10 (1999) 626–634.
- [20] J.-F. Cardoso, A. Souloumiac, Blind beamforming for non-gaussian signals, in: *IEE Proceedings F (Radar and Signal Processing)*, volume 140, IET, pp. 362–370.
- [21] A. Belouchrani, K. Abed-Meraim, J. Cardoso, E. Moulines, Second-order blind separation of temporally correlated sources, in: *Proc. Int. Conf. Digital Signal Processing*, Citeseer, pp. 346–351.
- [22] A. Delorme, J. Palmer, J. Onton, R. Oostenveld, S. Makeig, et al., Independent eeg sources are dipolar, *PloS one* 7 (2012) e30135.
- [23] H. Shen, S. Jegelka, A. Gretton, Fast kernel-based independent component analysis, *Signal Processing, IEEE Transactions on* 57 (2009) 3498–3511.
- [24] W. Kong, C. R. Vanderburg, H. Gunshin, J. T. Rogers, X. Huang, A review of independent component analysis application to microarray gene expression data, *Biotechniques* 45 (2008) 501.

- [25] J.-F. Cardoso, High-order contrasts for independent component analysis, *Neural computation* 11 (1999) 157–192.
- [26] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, *Neural networks* 13 (2000) 411–430.
- [27] D. Langlois, S. Chartier, D. Gosselin, An introduction to independent component analysis: Infomax and fastica algorithms, *Tutorials in Quantitative Methods for Psychology* 6 (2010) 31–38.
- [28] S. Choi, A. Cichocki, H.-M. Park, S.-Y. Lee, Blind source separation and independent component analysis: A review, *Neural Information Processing-Letters and Reviews* 6 (2005).
- [29] A. Cichocki, S.-i. Amari, Adaptive blind signal and image processing: learning algorithms and applications, volume 1, John Wiley & Sons, 2002.
- [30] A. Delorme, S. Makeig, Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis, *Journal of neuroscience methods* 134 (2004) 9–21.
- [31] T. M. Cover, J. A. Thomas, Elements of information theory, John Wiley & Sons, 2012.
- [32] A. Papoulis, S. U. Pillai, Probability, random variables, and stochastic processes, Tata McGraw-Hill Education, 2002.